

성공하는 빅데이터 거버넌스 및 모델링 전략

신정길

2014.9.30

contents

I. Big Data governance

II. Data Governance

III. Efficient Data Modeling

IV. Wrap up

contents

I. Big Data governance

II. Data Governance

III. Efficient Data Modeling

IV. Wrap up

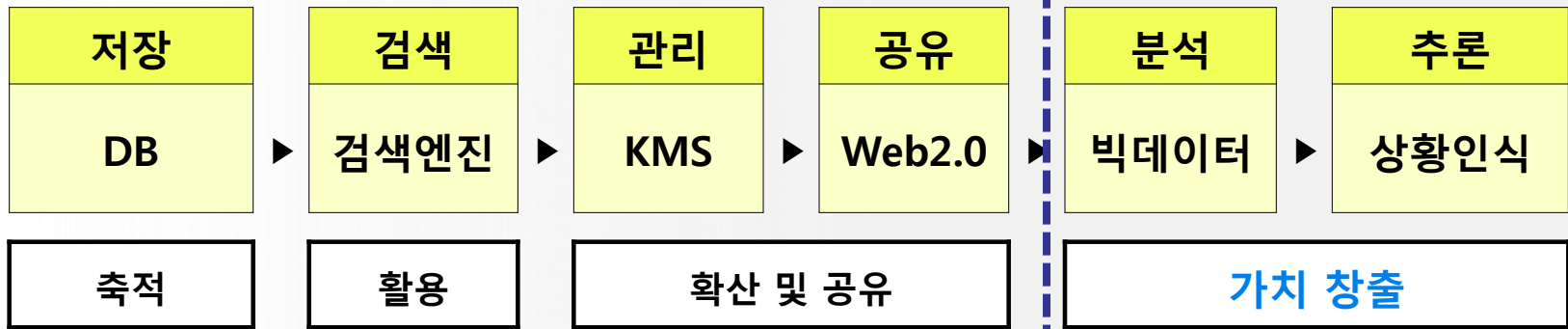
***Big Data**

Why?

Insight!

데이터의 과거, 현재, 미래

데이터의 과거-현재-미래



< 사람을 이긴 컴퓨터 '왓슨'의 힘은 '분석능력' >



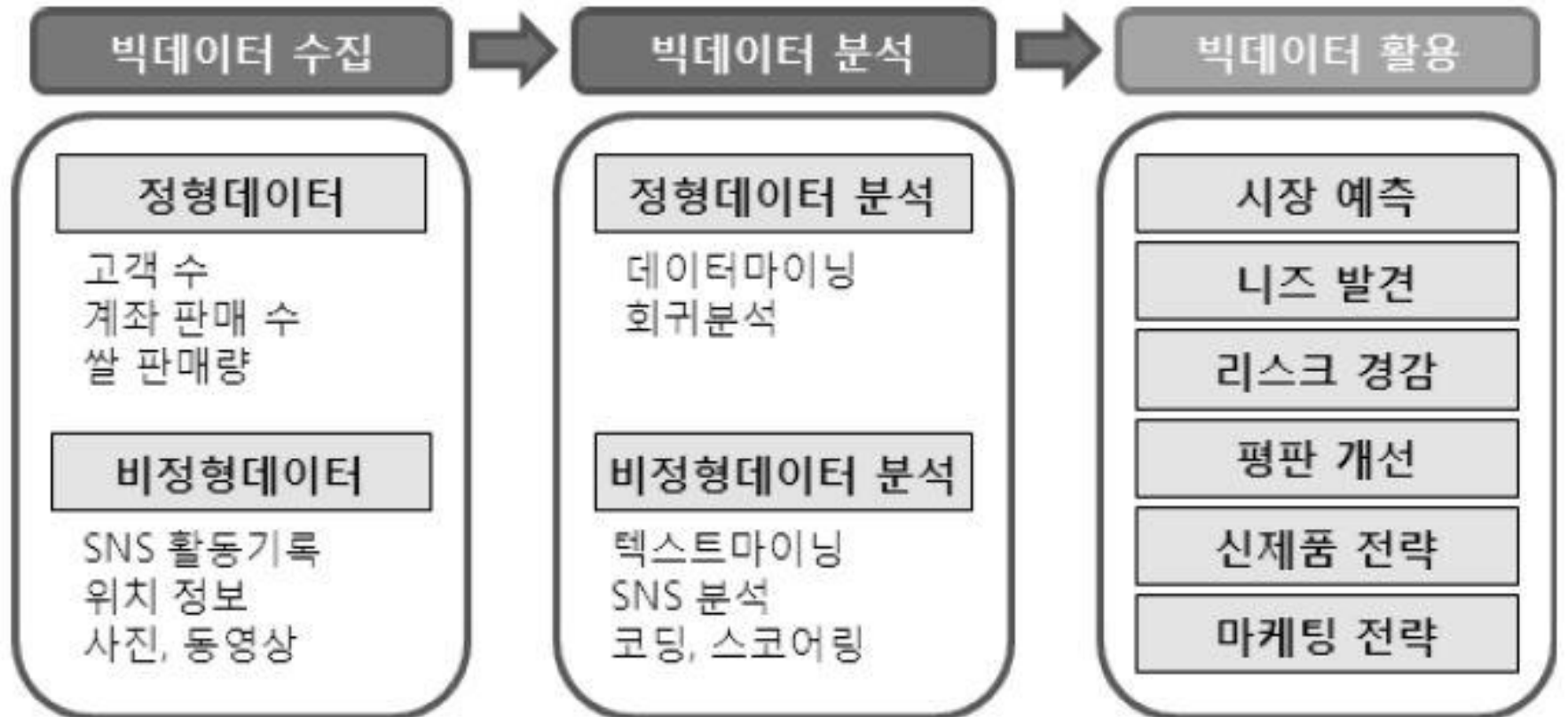
2011. 2 '제퍼디(Jeopardy!)' 최종전
IBM 슈퍼컴퓨터 '왓슨(Watson)' 압승



인공지능 슈퍼컴퓨터 '왓슨',
의료계와 금융계에 적용

빅데이터 활용

<빅데이터의 활용 프로세스>



빅데이터 생성 원천 ?

비즈니스 데이터

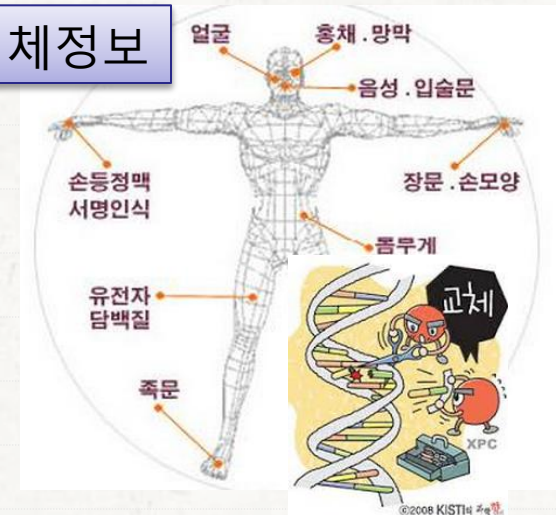
Tesco : 매달 15억 건 이상의 고객 관련 데이터를 수집



Social Data

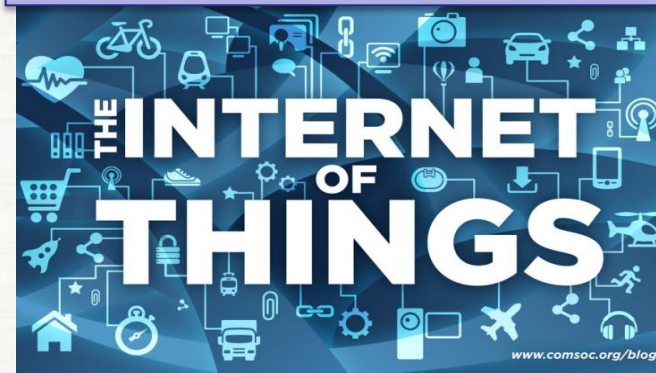


생체정보



사물인터넷

2014년 500억개 무선단말기(Ericsson)

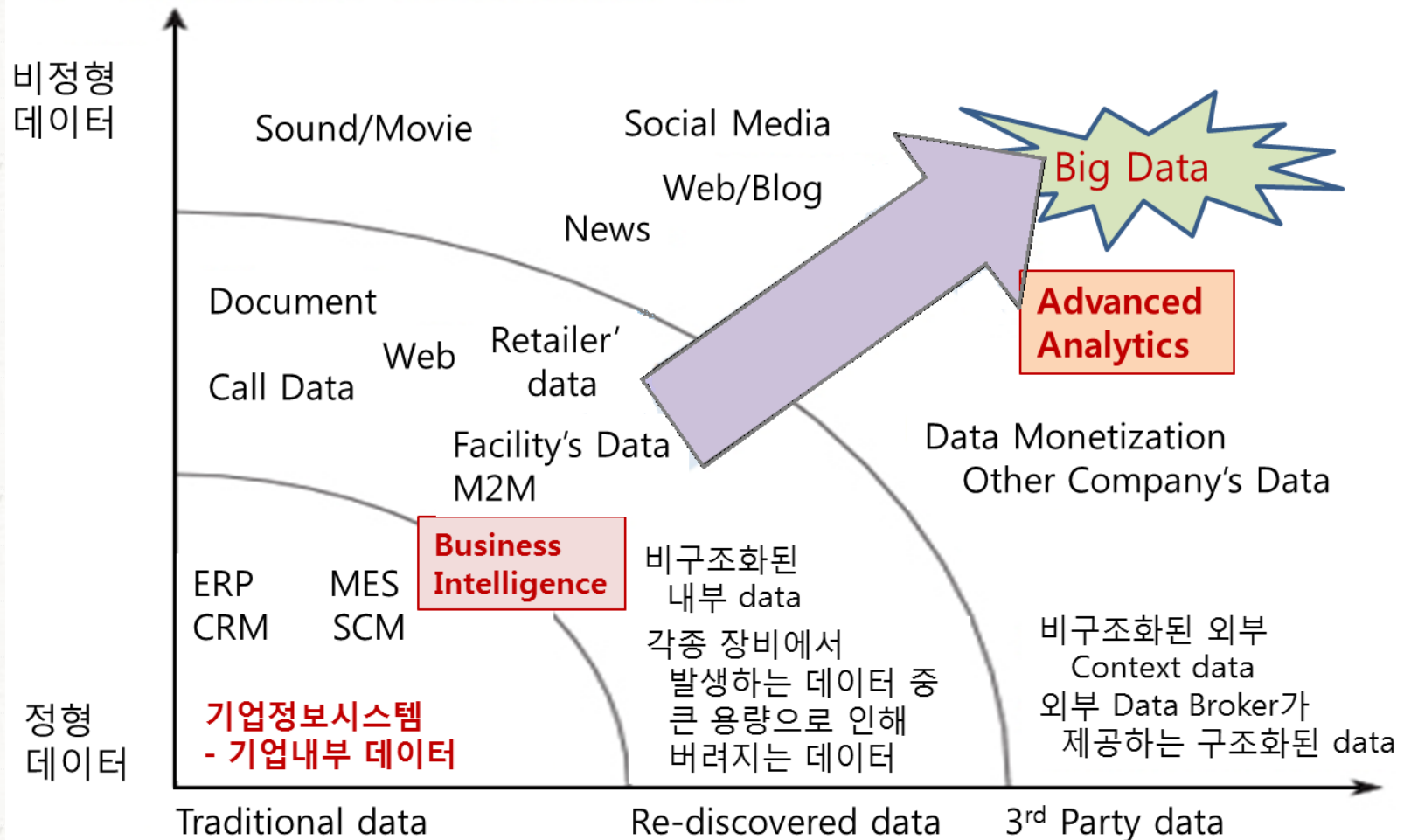


YouTube, CC 카메라 의료장비 동영상



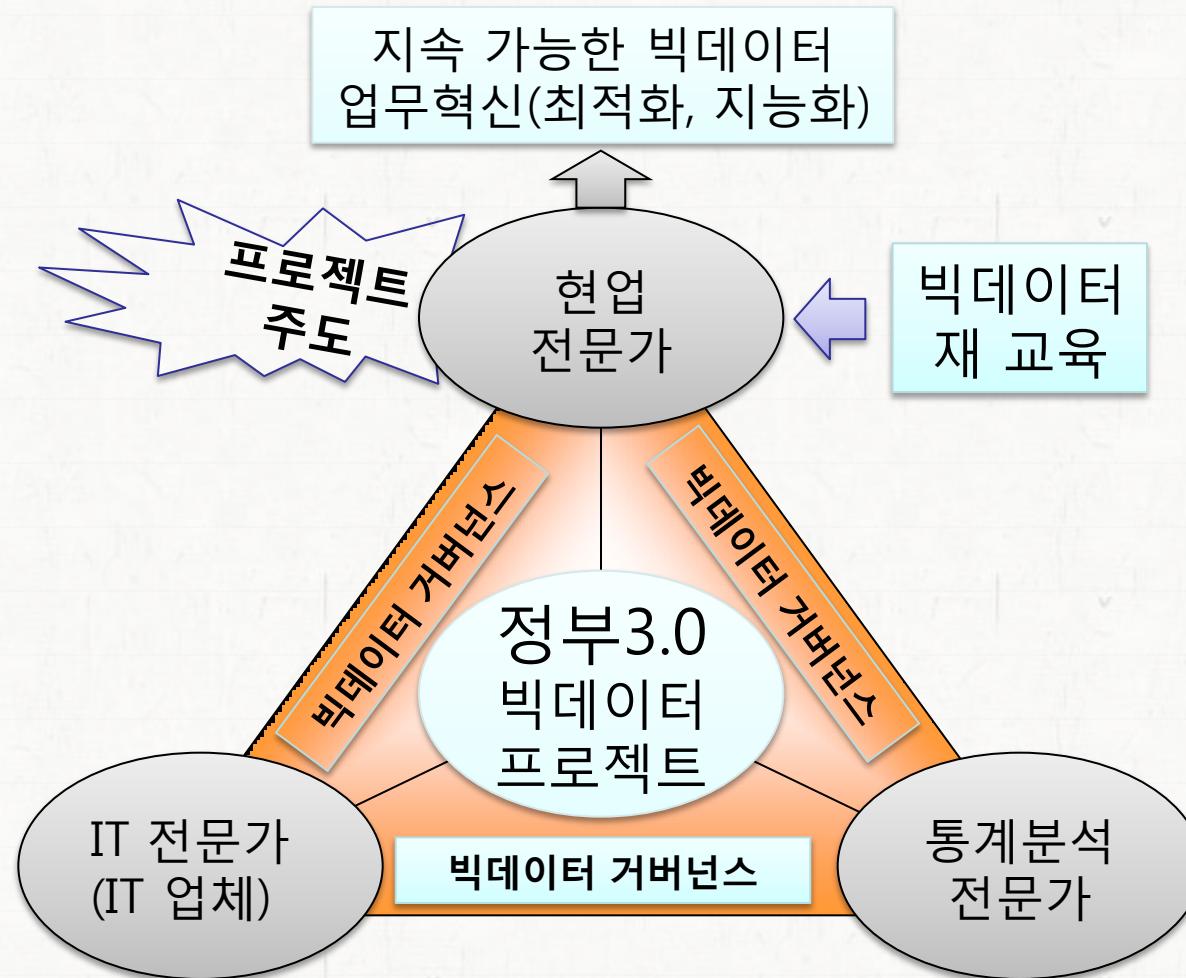
빅데이터 활용

• 분석 대상의 확대



빅데이터 활용과 거버넌스

• 성공적인 빅데이터 프로젝트 수행



빅데이터 거버넌스

- 데이터의 품질보장, 프라이버시 보호,
데이터 수명관리, 전담조직과 규정정립, 데이터
소유권과 관리권 명확화 등을 통하여
- 빅데이터가 적시에 필요한 사람에게 제공되도록
체계를 확립하는 것

빅데이터 거버넌스

빅데이터 거버넌스가 확립되지 못하면 :

- 품질 낮은 데이터를 중요한 의사결정에 사용함으로써 심각한 문제 야기
- 개인 프라이버시 관련 데이터로 인해 빅브라더의 우려 현실화
- 폭증하는 데이터의 수명관리 소홀로 IT 비용 급증
- 데이터 관련 문제를 전담하고 책임지는 조직과 인력이 없다면 빅데이터 효과는 일회성에 그칠 것임

빅데이터 거버넌스 사례

• 과학기술분야 - 데이터 품질



1999년 예정된 궤도보다 170km 아래에 진입, 폭발
미국-영국 과학자들이 (복잡한 수식에서)
수치의 단위를 **newton 대신에 pound**로 사용한 것이 원인



9개월간의 항해
3억 3000만달러 손실
우주산업 후퇴



리스크, 재난, 의료 등 다양한 빅데이터 분야에서 복잡하고 방대하며, 실시간으로 발생하는 데이터를 오해하거나 잘못 적용하여 오판을 일으킨다면?

: 빅데이터 거버넌스의 중요성

빅데이터 거버넌스 프레임워크

• IT & Data 거버넌스

IT 거버넌스



Data 거버넌스

정형

3Vs

BigData 거버넌스

3Vs로 더욱 중요

IT 자원의 관리 체계
- HW, SW, 인력, 제도
- Data ?

효과성: IT가 전사적 목표 달성에 기여?
투명성: IT활동이 원칙과 기준에 따라 수행?
책임성: IT성과에 대한 책임자?

전사적 데이터에 관하여 관리 정책, 지침, 표준, 전략 및 방향을 수립하고, 관리할 수 있는 조직과 인력을 배치하는 데이터 관점에서의 IT 관리 체계

데이터의 통합/정제
프라이버시
품질/표준화
MDM
수명주기
지침과 규정

빅데이터 거버넌스 프레임워크

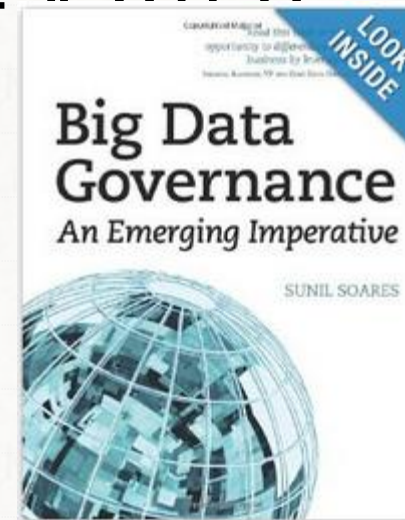
• 빅데이터 거버넌스의 특징

- 3Vs 특징으로 인해 BDG가 더욱 어렵고 중요함
 - Volume - 수명주기 관리로 비용절감 필요성
 - Variety - Metadata 관리, 품질관리, 프라이버시 보호의 중요성
 - Velocity - 더욱 빠른 통제 및 관리 체계 필요
- 빅데이터에 대한 거버넌스는 현재 정립 중

데이터 거버넌스

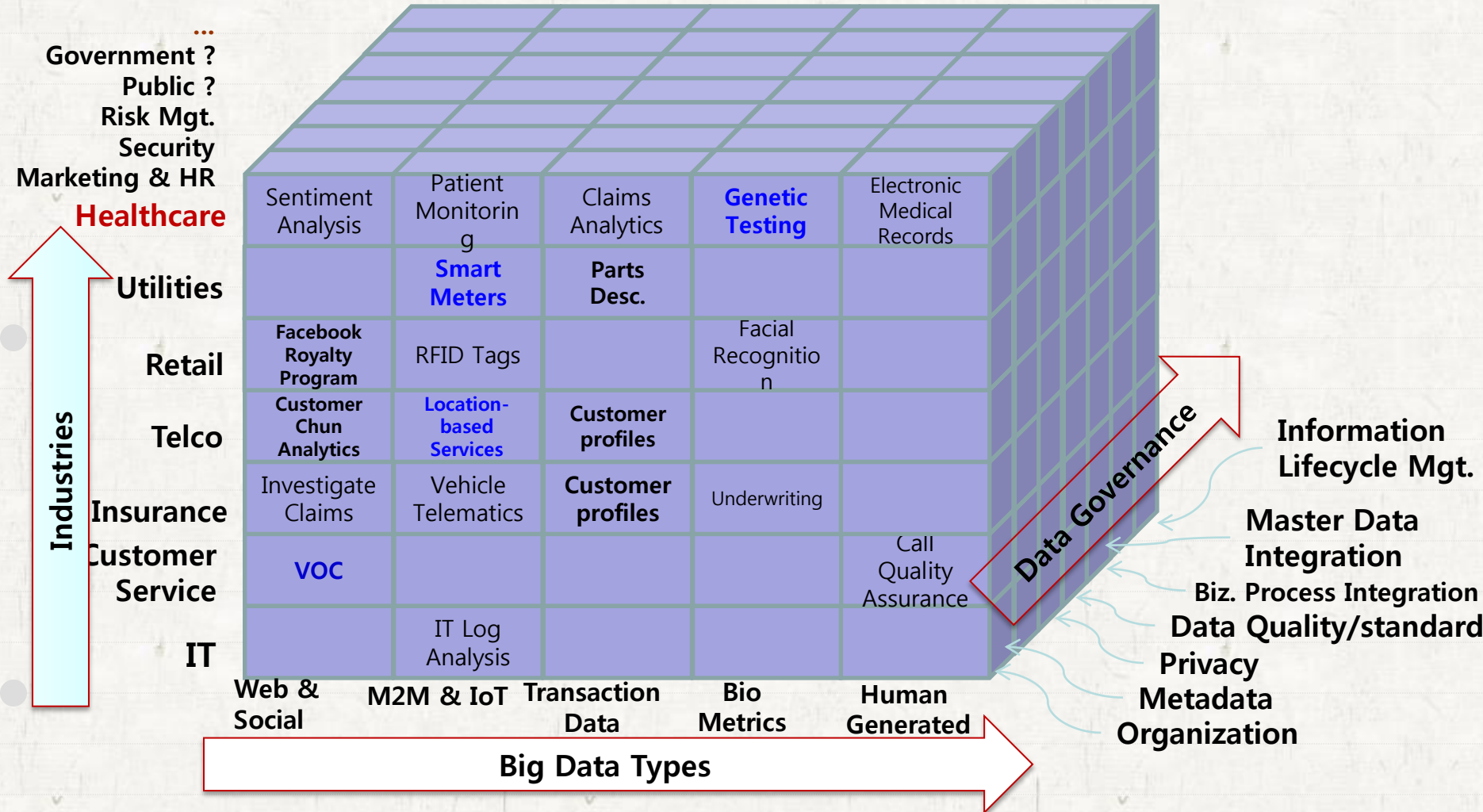
Data 품질과 표준화
 프라이버시 보호
 메타 데이터 관리
 데이터 수명관리
 관련 인력, 조직, 규정

+ 빅데이터
 3Vs =



빅데이터 거버넌스 프레임워크

BDG는 산업/데이터유형/데이터거버넌스 세 가지 측면에서 관리해야 함



빅데이터 유형과 거버넌스

• 다양한 유형의 데이터가 생성됨

Web and Social Media

- Clickstream Data
- Twitter Feeds
- Facebook Postings
- Web Content

카카오톡
블로그, 상품게시판 등

Biometrics

- Facial Recognition
- Genetics

홍채, 지문, 필체, ...

Machine-to-Machine

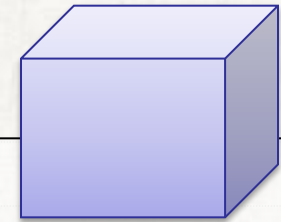
- Utility Smart Meter Readings
- RFID Readings
- Oil Rig Sensor Readings
- GPS Signals

Human Generated

- Call Center Voice Recordings
- Email
- Electronic Medical Records

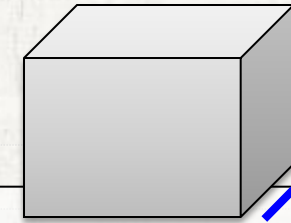
Big Transaction Data

- Healthcare Claims
- Telecommunications Call Detail Records
- Utility Billing Records



Big data type

빅데이터 거버넌스 항목



데이터
거버넌스
항목

• 조직 (Organization)

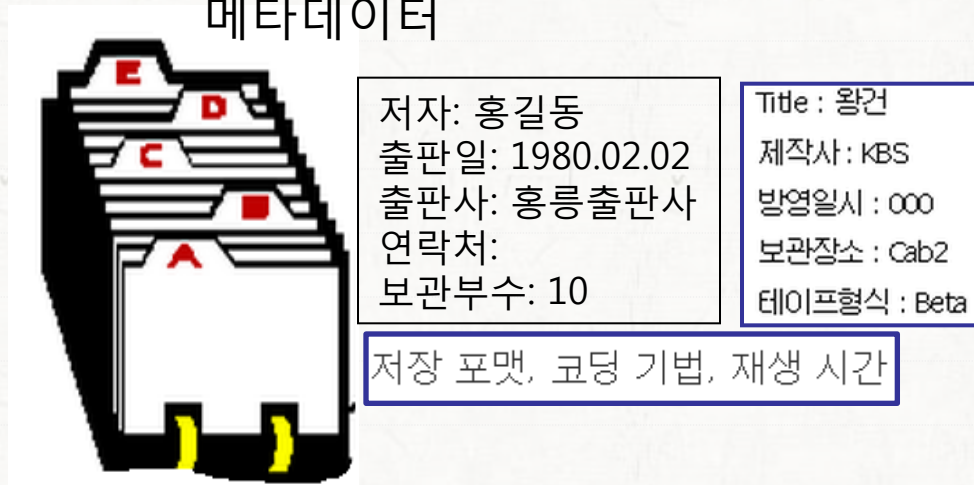
- 데이터 거버넌스 위원회에 데이터 과학자 추가
- 기존 데이터 관리자에게 추가적인 책임을 할당
 - 예를 들어, 고객 데이터 관리자에게 마스터 데이터 리파지토리 안에서 페이스북, 블로그, 트위터 계정에 대한 책임을 부과함

빅데이터 거버넌스 항목

• 메타 데이터 관리

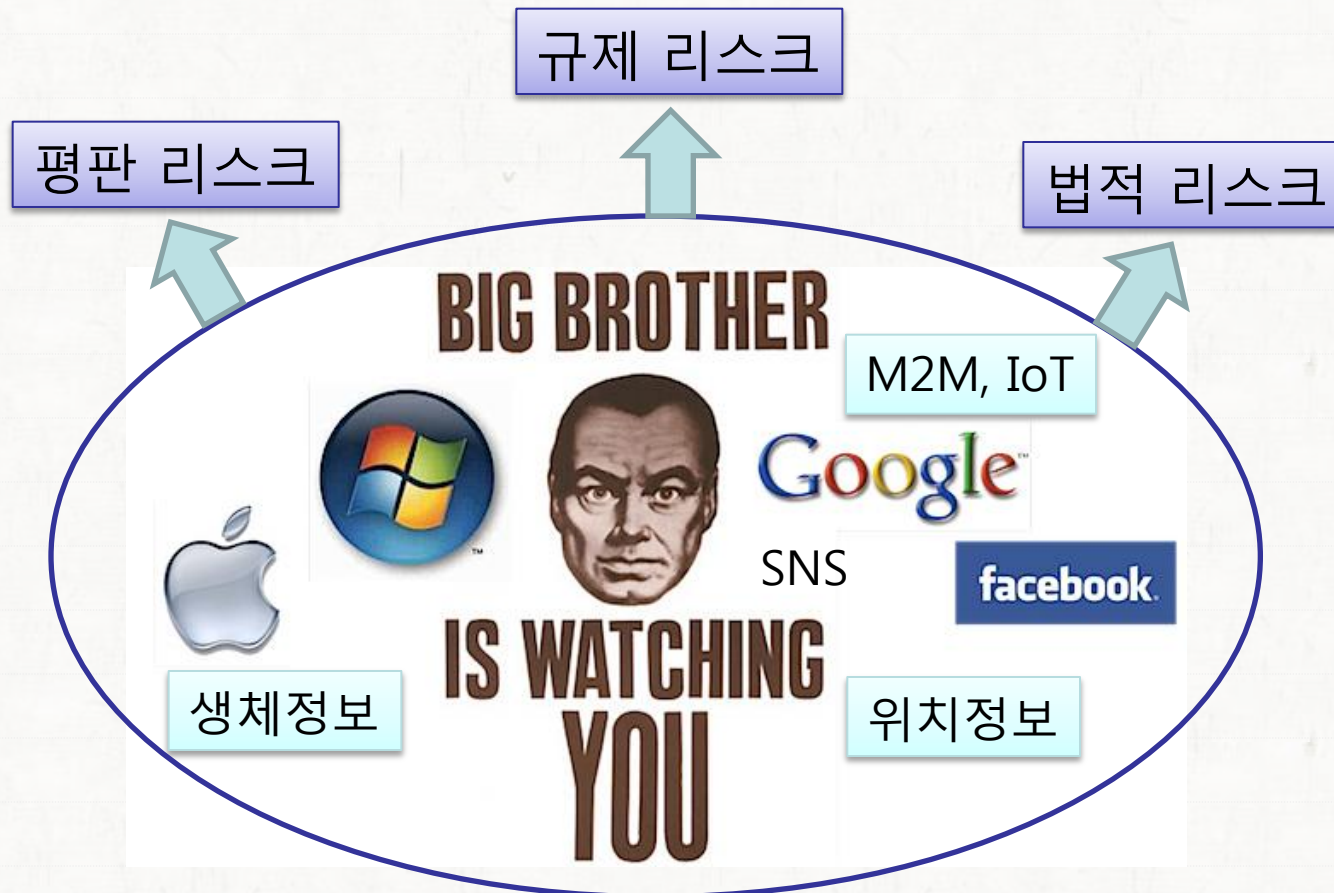
- 빅데이터의 종류를 파악하여 목록을 작성하고 관리함
- 다양한 빅데이터의 소스, 추출-정제-변환-통합규칙, 저장 형식, 관리자, 관련 용어 등 관리 대상을 정하고, 표준화함

메타데이터



빅데이터 거버넌스 항목

• 프라이버시 보호 중요성 증대



빅데이터 거버넌스 항목

• 데이터 품질 관리

- 데이터에 관한 품질측정, 품질개선, 품질인증과 무결성 보장, 표준화 등의 규정을 정하는 것
- 데이터 품질저하로 인한 문제 사례
 - 금융계좌 주민번호 398만개 오류로 이자소득 금융과세 못해 (조선일보, 2004년 1월 4일)
- 빅데이터의 활성화와 데이터 품질 문제
 - 재난, 위험관리, 의료 등에서 저 품질로 인한 문제 발생 가능성
- 미국의 경우 Data Quality Act(데이터 품질법) 제정
 - 공공목적으로 배포되는 일체의 데이터는 신뢰성, 안전성, 정확성, 이용 편의성을 일정 수준 이상으로 유지해야 함

빅데이터 거버넌스 항목

• 정보 수명주기 관리

- 방대한 데이터를 운영 시스템이나 분석 시스템에 언제까지 유지할 것인지를 결정하고, 나머지는 보관소(archive)에 저장하거나 폐기하도록 해야 함 : IT 비용 절감 & 위험감소

데이터 수명관리



생체정보

트랜잭션 data

M2M | IoT

• 지속 가능한 빅데이터가 되려면 ?

- 빅데이터의 활용보다 데이터 거버넌스가 먼저

빅데이터 거버넌스는

데이터의 품질보장, 프라이버시 보호, 데이터 수명관리, 전담조직과 규정 정립, 데이터 소유권과 관리권 명확화 등을 통하여 빅데이터가 적시에 필요한 사람에게 제공되도록 체계를 확립하는 것

빅데이터 거버넌스가 확립되지 못하면

- 품질이 낮은 데이터를 중요한 의사결정에 사용함으로써 심각한 문제를 야기
- 개인 프라이버시 관련 데이터로 인해 빅브라더의 우려가 현실화
- 폭증하는 데이터의 수명관리 소홀로 인해 IT 비용이 급증
- 데이터 관련 문제를 전담하고 책임지는 조직과 인력이 없다면 빅데이터 효과는 일회성에 그칠 것임

contents

I. Big Data governance

II. Data Governance

III. Efficient Data Modeling

IV. Wrap up

*Data Governance

Why?

IT 환경에 대한 도전

IT Challenges

Evolving Database Ecosystems



Volume, Velocity, Variety



Agile Development Cycles



Compliance



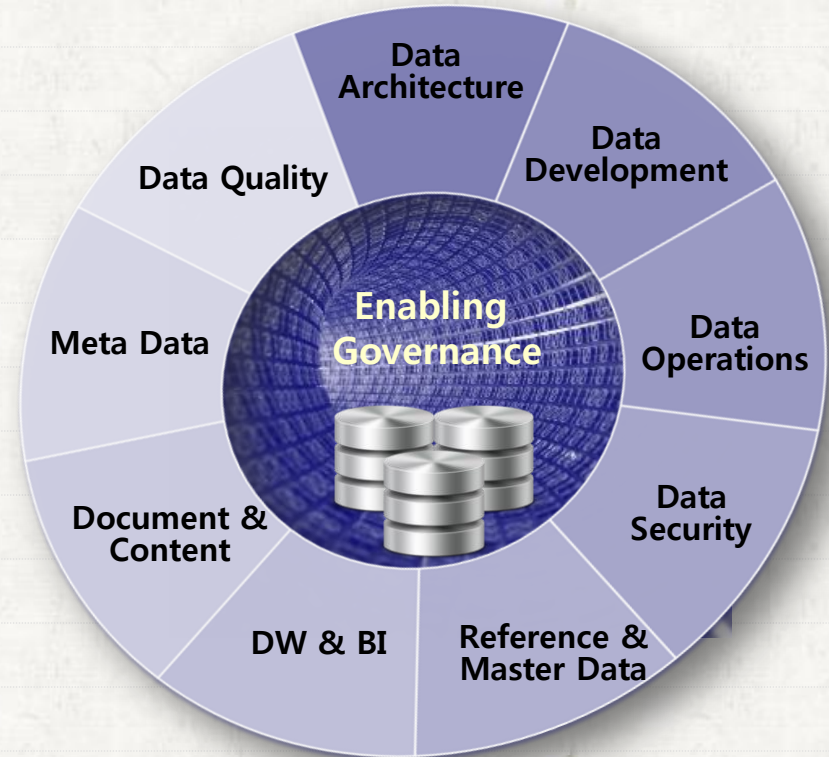
Organizational Transparency



Maximizing IT Infrastructure



- 빠른 DB 환경 변화
- 폭발적으로 증가하는 데이터
- 빈번한 요구사항 변경 및 개발
- 정보(데이터)의 중요성 부각
- IT 중요성 및 투자 증대



출처: Data Management International (www.dama.org)

DATA Governance

데이터 사용



- 지식 근로자가 정보를 조회하고 획득하는 시간 단축

데이터 생산



- 중복 작업 제거
- 중복되거나 불필요한 인프라 통합 및 응용프로그램 제거

데이터 품질



- 재작업 제거 및 경영 합리화
- 분석 및 리포팅 오류 감소

운영효율성



- 향상된 캠페인 타게팅
- 정확한 제안 및 프라이싱
- 변화에 대한 향상된 대응

Data Management Challenges

Going beyond native database tools

Embarcadero

Data Governance enabling cross-platform database tools

비즈니스와 데이터 연결

- ✓ DB 관리 생태계
- ✓ 메타데이터 리파지토리
- ✓ 아키텍처/모델링툴 통합

툴 표준화

- ✓ 크로스플랫폼 기능
- ✓ 버전 컨트롤 통합

생산성 향상

- ✓ 향상된 분석 통계와 뷰
- ✓ 상세한 비교 및 롤백 기능

Vendor Tools

Productive database tools for your database platform

Database Administration

Performance Tuning

Change Management

Evolving Database Ecosystems



Volume, Velocity, Variety



Agile Development Cycles



Compliance



Organizational Transparency



Maximizing IT Infrastructure



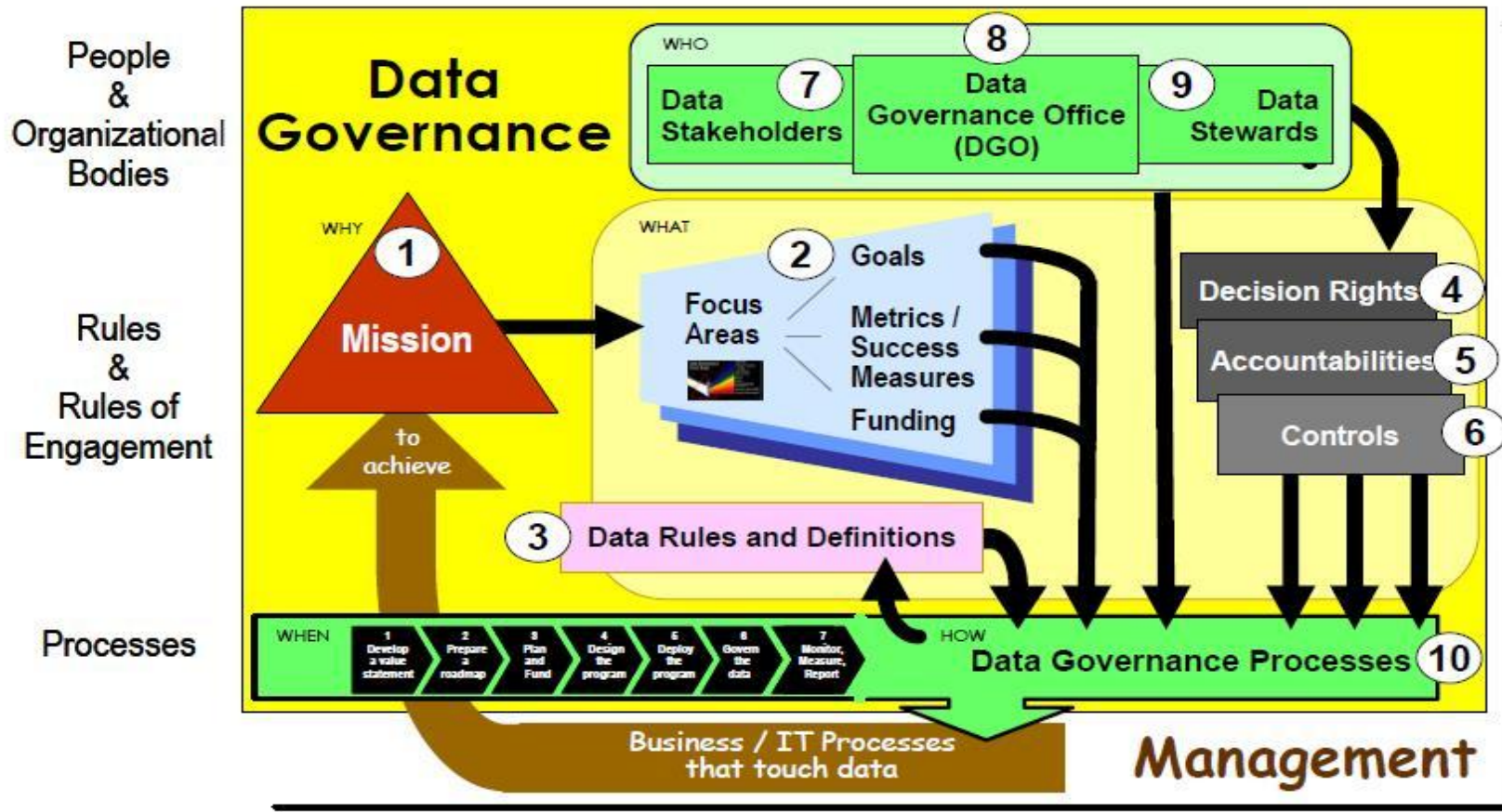
Data Governance



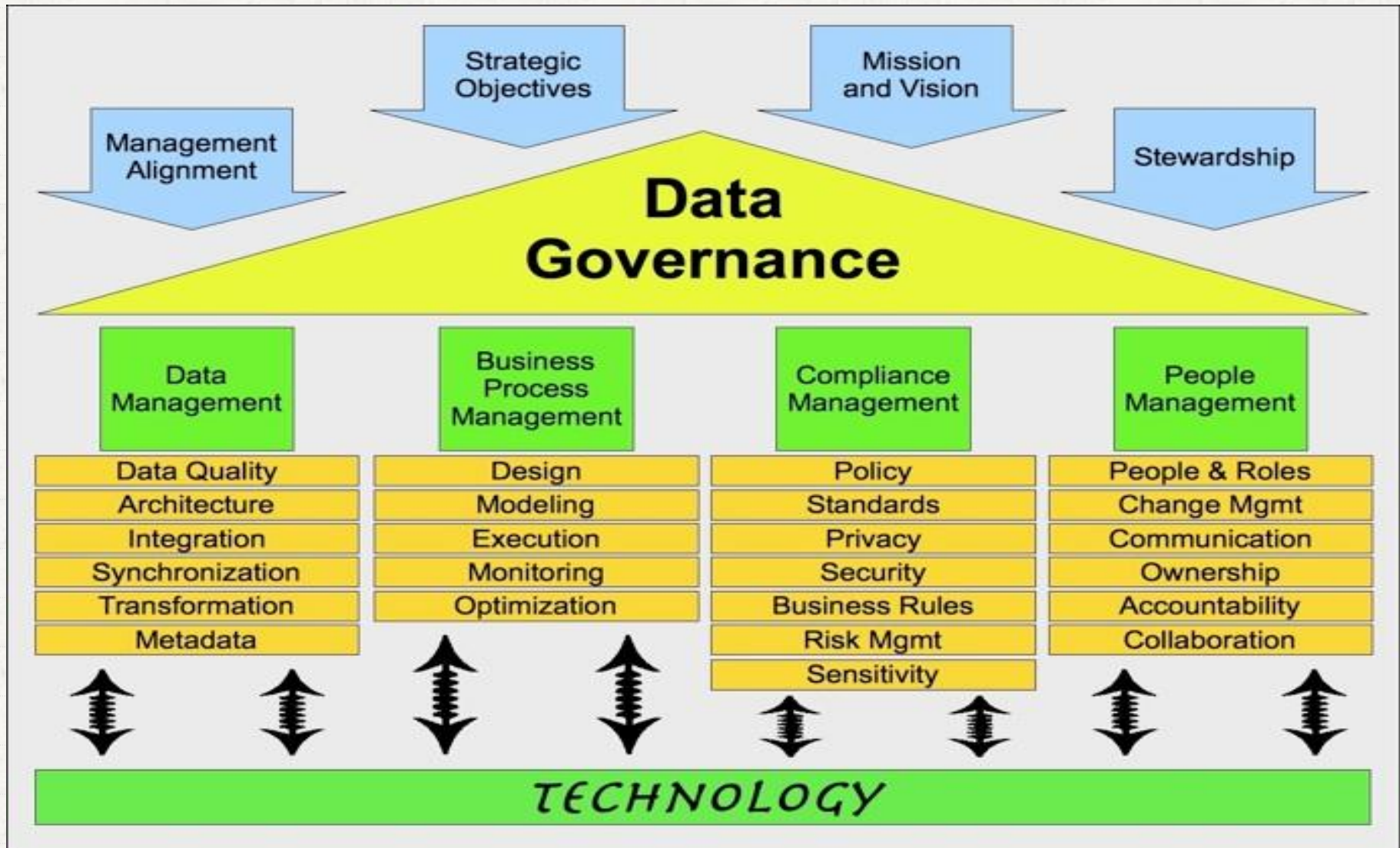
The DGI Data Governance Framework

from The Data Governance Institute

10 Components of a Data Governance Program



Data Governance



contents

I. Big Data governance

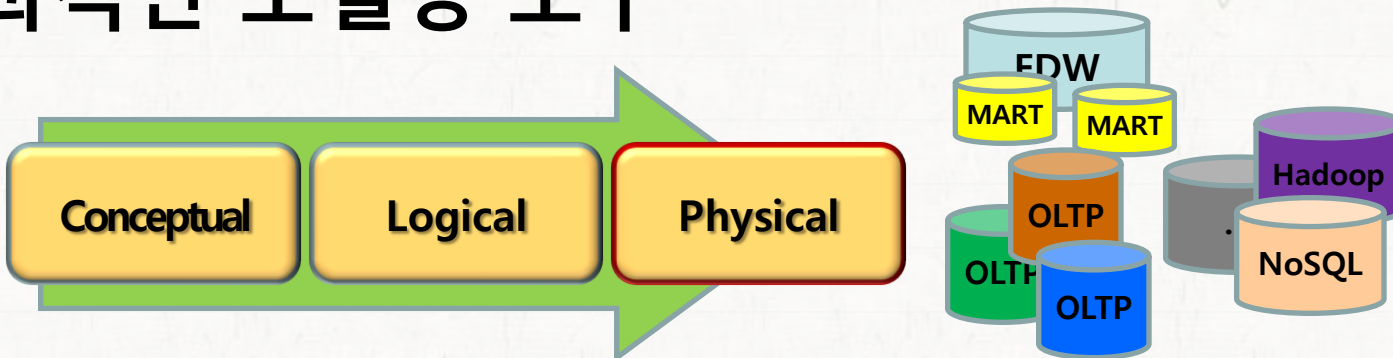
II. Data Governance

III. Efficient Data Modeling

IV. Wrap up

Efficient Data Modeling

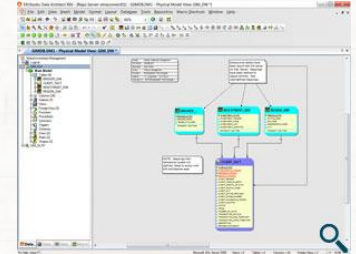
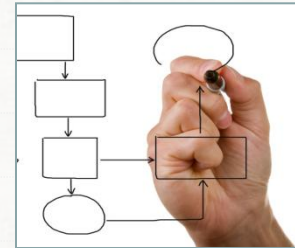
- **Utility of Data Modeling?**
- **Relational vs. NoSQL**
 - 객체 모델 지향 vs. 쿼리결과지향
 - 정규화 vs. 비정규화
 - HW 리소스 절감 vs. 쿼리/질의/읽기 최적화
- **물리모델링 - DBMS에 대한 지식 필요**
- **효과적인 모델링 도구**



Efficient Data Modeling

• Utility of Data Modeling?

- 다양한 DBMS 탄생
- 수많은 응용시스템
- 빈번한 요구사항 변경
- 데이터 중요성 부각
- 협업 필요



• 툴의 기능 확대(데이터 통찰력)

- 다양한 DB 객체 설계
- 데이터의 생성, 변경, 흐름 관리
- 다양한 DBMS 지원
- 협업 기능 강화



Efficient Data Modeling

• 주요 모델링 툴

구분	제조사	제품명
국외	CA Technologies	Erwin
	Embarcadero Technologies	ER/Studio
	SAP Sybase	PowerDesigner
국내	엔코아	DA#
	토마토시스템	exERD

Source : KDB 2014 데이터베이스 백서

“기술지원”

“편의성”

“표준표기법”

“예산”

“지원DBMS”

“모델링기능”

“사용용도”

“제품신뢰성”

“업데이트”

ER/Studio XE6

• 엔터프라이즈급 데이터모델링

– DB 환경 변화에 대응하는 빠른 업데이트

DBMS 지원

- DBMS 환경 변화(신규 버전, 새로운 DB 등)에 빠른 대응
- Firebird, Interbase, Greenplum, Hitachi HiRDB
- IBM DB2, Informix Online and SE, Informix dynamic server
- MS Access, **MS SQL Server 2014**
- **Oracle 12c**, Sybase ASE / ASA / IQ / Watcom SQL
- Teradata, MySQL, **PostgreSQL**
- **Netezza, Hadoop Hive, Mongo DB, Cloudera** 등

ER/Studio XE6

• 엔터프라이즈급 데이터모델링

- 데이터의 생성과 소멸까지(데이터 통찰력 제공)

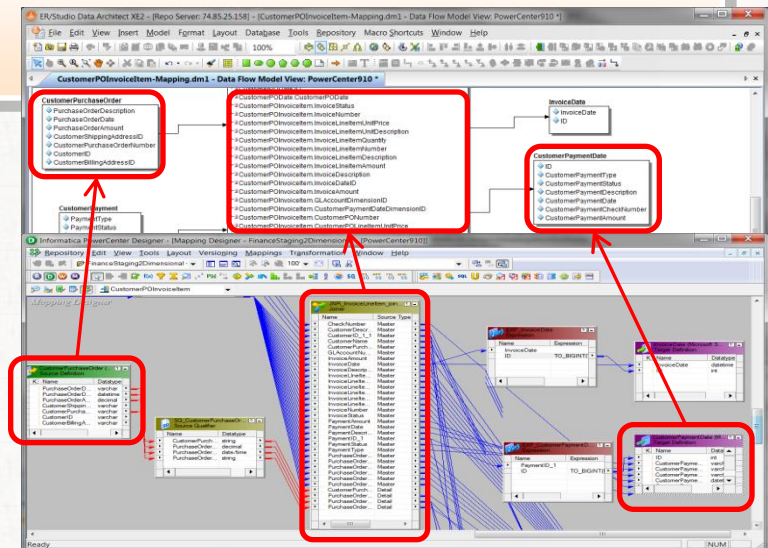
데이터 통찰력

- Data Lineage 시각화 및 탐색(주요 ETL 툴 Import/Export)
 - Informatica Power Center
 - SAB Business Objects Data Integrator
 - IBM Data Stage
 - Oracle ODI and OWB
 - Microsoft SSIS 등

ETL 시각지대 제거

노동집약적인 문서작업 자동화

순쉬운 데이터거버넌스 리포팅



ER/Studio XE6

• 엔터프라이즈급 데이터모델링

- 현 시대 IT 요구사항 반영

기업 IT
환경 반영

- 단일 논리모델에 여러 가지 물리 모델 생성
 - 예 : 논리모델 - 인사(Oracle용 물리모델), 회계(Sybase용 물리모델) 등
- 다단계 생성 가능한 서브모델
 - 예 : 전사모델 - 학사행정SubModel(입학sub, 등록sub ...)
- 다양한 매크로 및 API
 - 범용 언어를 이용한 매크로(SAX BASIC)
 - 다양한 API 제공으로 타 응용프로그램에서 손쉬운 연동
 - 다양한 자동화 기능 구현으로 생산성 극대화

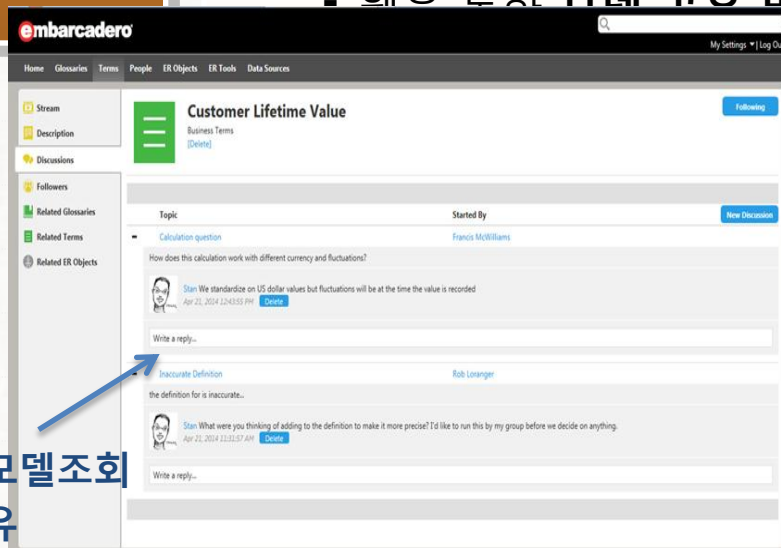
ER/Studio XE6

- 엔터프라이즈급 데이터모델링
 - 데이터 모델/표준용어에 대한 협업 강화

협업

- Repository
 - Object 단위 Check In/Out으로 다수 유저가 동시 작업
 - 다양한 권한 관리
- Team Server

웹을 통한 모델 공유 및 비즈니스 용어 관리
 시간 협업 활성화
 더 연동



웹을 통한 모델조회
 및 의견 공유

```

1 SELECT
2   EMPLOYEES.FIRST_NAME,
3   JOBS.JOB_TITLE,
4   EMPLOYEES.SALARY,
5   EMPLOYEES.Monthly salary of the employee in Euros.
6 FROM
7   HR.EMPLOYEES,
8   HR.JOBS
9  WHERE
10  HR.EMPLOYEES.JOB_ID = HR.JOBS.JOB_ID
  
```

쿼리 작성 시
 비즈니스 용어 제공

ER/Studio XE6

• 엔터프라이즈급 데이터모델링

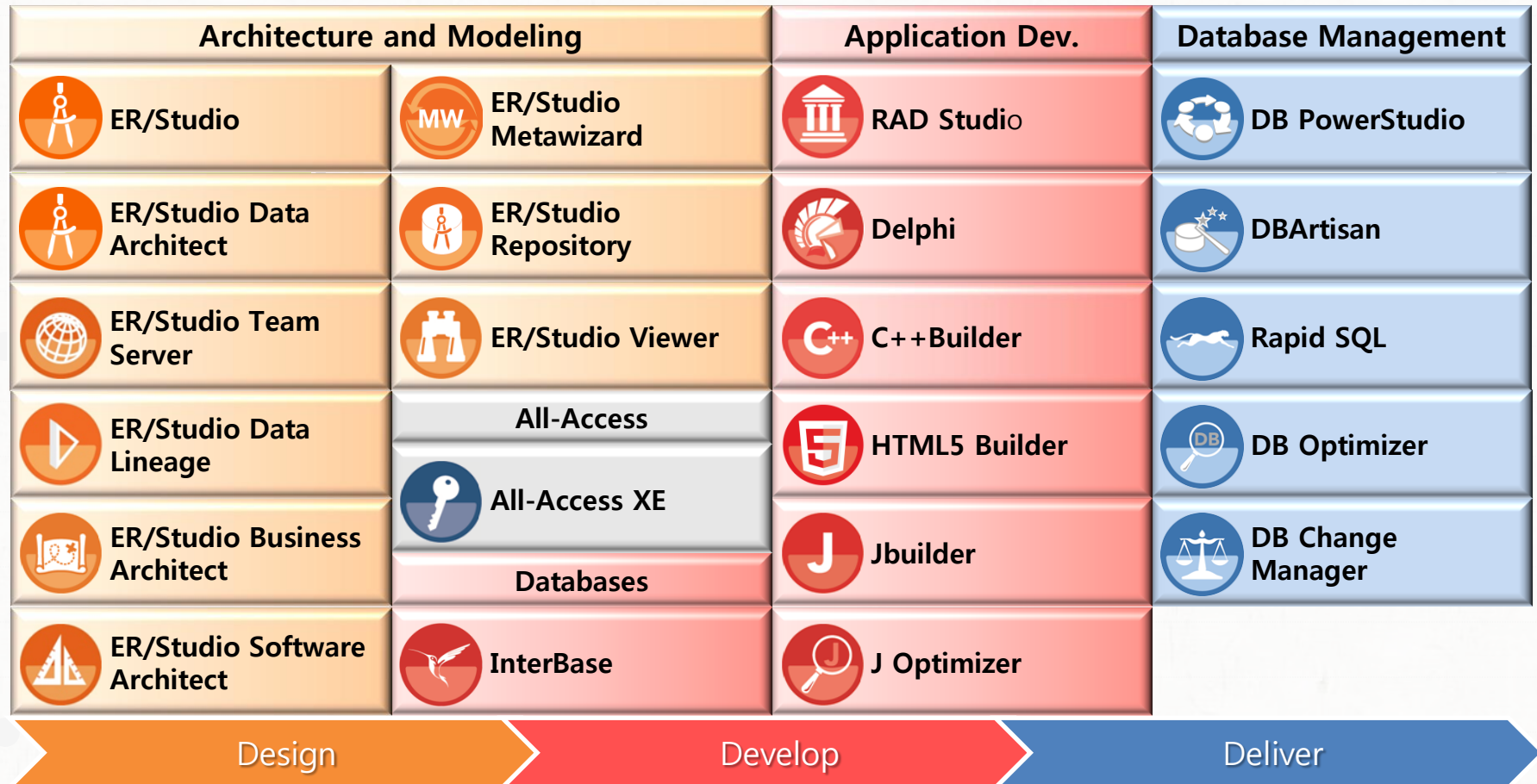
- 데이터의 생성과 소멸까지(데이터 통찰력 제공)

신뢰성

- 엠바카데로의 주력 솔루션으로 지속적인 발전
- 본사의 강력한 기술지원 조직
- 빠른 업데이트(매년 최소 1회 Major 업그레이드)
- DB설계/개발에서 부터 응용프로그램 개발까지 Line Up

ER/Studio XE6

• 엔터프라이즈급 데이터모델링



Heterogeneous, cross-platform tools to design, develop, and deliver your data and applications

A scenic autumn landscape featuring a stone path winding through a forest of trees with vibrant red, orange, and yellow foliage. The ground is covered in fallen yellow leaves. In the foreground, there are several dark green bushes and a wooden bench partially visible on the right. The background shows a hillside with more trees and a clear sky.

Wrap uP !

질문 ?



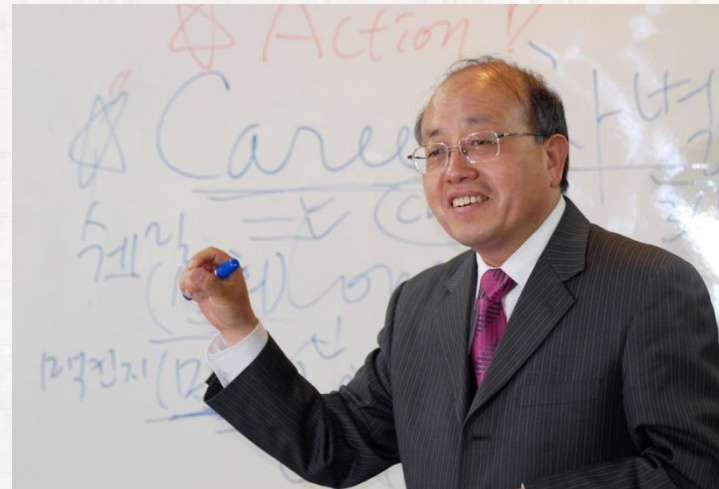


감사합니다 !!!

신정길



- ❖ 02-595-4288
- ❖ ask@embarcadero.kr



- ❖ 010-3294-3829
- ❖ jingles@naver.com